UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
10/626,875	07/25/2003	Francine R. Chen	D/A3053/311291	1845

52725 7590 03/08/2006

CHRISTIAN AUSTIN-HOLLANDS ESQ.
P.O. BOX 170325
SAN FRANCISCO, CA 94117

EXAMINER	
PONIKIEWSKI, TOMASZ	
ART UNIT	**PAPER NUMBER**
2165	

DATE MAILED: 03/08/2006

Please find below and/or attached an Office communication concerning this application or proceeding.

Office Action Summary	Application No.	Applicant(s)	
	10/626,875	CHEN ET AL.	
	Examiner	Art Unit	
	Tomasz Ponikiewski	2165	

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTH(S) OR THIRTY (30) DAYS, WHICHEVER IS LONGER, FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

1)☐ Responsive to communication(s) filed on _____.

2a)☐ This action is **FINAL**. 2b)☒ This action is non-final.

3)☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims

4)☒ Claim(s) 1-88 is/are pending in the application.

4a) Of the above claim(s) _____ is/are withdrawn from consideration.

5)☐ Claim(s) _____ is/are allowed.

6)☒ Claim(s) 1-88 is/are rejected.

7)☐ Claim(s) _____ is/are objected to.

8)☐ Claim(s) _____ are subject to restriction and/or election requirement.

Application Papers

9)☒ The specification is objected to by the Examiner.

10)☐ The drawing(s) filed on _____ is/are: a)☐ accepted or b)☐ objected to by the Examiner.

Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).

Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).

11)☐ The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

Priority under 35 U.S.C. § 119

12)☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).

a)☐ All b)☐ Some * c)☐ None of:

1.☐ Certified copies of the priority documents have been received.

2.☐ Certified copies of the priority documents have been received in Application No. _____.

3.☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

* See the attached detailed Office action for a list of the certified copies not received.

Attachment(s)

1) ☒ Notice of References Cited (PTO-892)
2) ☐ Notice of Draftsperson's Patent Drawing Review (PTO-948)
3) ☒ Information Disclosure Statement(s) (PTO-1449 or PTO/SB/08) Paper No(s)/Mail Date 10/28/2005, 04/20/2004, 07/25/2003
4) ☐ Interview Summary (PTO-413) Paper No(s)/Mail Date. _____.
5) ☐ Notice of Informal Patent Application (PTO-152)
6) ☐ Other: _____.

DETAILED ACTION

1. Claims 1-88 are pending.

Specification

2. The disclosure is objected to because of the following informalities: In the section titled Incorporation by Reference application serial number for "Systems and Methods for New Event Detection" is missing.

Appropriate correction is required.

3. The disclosure is objected to because of following informalities: in the section titled Detailed Description of Exemplary Embodiments on page 34 in paragraph 0129 the recitation "later developed" draws toward the unpredictability of the future. This is not clear and concise language.

Claim Objections

4. Claim 83 is objected to because of the following informalities: in line 27 the word "the" is written twice, one should be deleted. Appropriate correction is required.

Double Patenting

5. The nonstatutory double patenting rejection is based on a judicially created doctrine grounded in public policy (a policy reflected in the statute) so as to prevent the unjustified or improper timewise extension of the "right to exclude" granted by a patent and to prevent possible harassment by multiple assignees. A nonstatutory obviousness-type double patenting rejection is appropriate where the conflicting claims are not identical, but at least one examined application claim is not patentably distinct from the reference claim(s)

because the examined application claim is either anticipated by, or would have been obvious over, the reference claim(s). See, e.g., In re Berg, 140 F.3d 1428, 46 USPQ2d 1226 (Fed. Cir. 1998); In re Goodman, 11 F.3d 1046, 29 USPQ2d 2010 (Fed. Cir. 1993); In re Longi, 759 F.2d 887, 225 USPQ 645 (Fed. Cir. 1985); In re Van Ornum, 686 F.2d 937, 214 USPQ 761 (CCPA 1982); In re Vogel, 422 F.2d 438, 164 USPQ 619 (CCPA 1970); and In re Thorington, 418 F.2d 528, 163 USPQ 644 (CCPA 1969).

A timely filed terminal disclaimer in compliance with 37 CFR 1.321(c) or 1.321(d) may be used to overcome an actual or provisional rejection based on a nonstatutory double patenting ground provided the conflicting application or patent either is shown to be commonly owned with this application, or claims an invention made as a result of activities undertaken within the scope of a joint research agreement.

Effective January 1, 1994, a registered attorney or agent of record may sign a terminal disclaimer. A terminal disclaimer signed by the assignee must fully comply with 37 CFR 3.73(b).

6. Claims 1, 20, 39, 58, 81-84 are provisionally rejected on the ground of nonstatutory obviousness-type double patenting as being unpatentable over claims 1, 16, 31-32 of copending Application No. 10/626,856. Although the conflicting claims are not identical, they are not patentably distinct from each other because the claims use determining steps that are clearly similar. For example in claim 1 of the instant application applicant states "determining source-identified training stories", in claim 1 of application 10/626,856 applicant states "determining a source-identified story corpus, each story associated with at least one event". In effect both claims state the same thing. Other steps in reminder of the claims follow the same reasoning.

This is a provisional obviousness-type double patenting rejection because the conflicting claims have not in fact been patented.

Claim Rejections - 35 USC § 112

7. The following is a quotation of the second paragraph of 35 U.S.C. 112:

> The specification shall conclude with one or more claims particularly pointing out and distinctly claiming the subject matter which the applicant regards as his invention.

8. Claims 11, 30, 49, and 68 are rejected under 35 U.S.C. 112, second paragraph, as being indefinite for failing to particularly point out and distinctly claim the subject matter which applicant regards as the invention.

9. Claims 11, 30, 49, and 68 recite the limitation "the inter-story similarity metrics" in the body of the claims. There is insufficient antecedent basis for this limitation in the claim.

Claim Rejections - 35 USC § 101

10. 35 U.S.C. 101 reads as follows:

> Whoever invents or discovers any new and useful process, machine, manufacture, or composition of matter, or any new and useful improvement thereof, may obtain a patent therefor, subject to the conditions and requirements of this title.

11. Claims 1,20, 39, 58, 77 and 81-84 are rejected under 35 U.S.C. 101 because the claimed invention is directed to non-statutory subject matter.

Claims 1, 39, 77, 82 and 84 list computational steps in a program without tangible, useful, concrete result. The claims do not list any hardware (i.e. computer) tied to the method steps in order to realize the "determining" functionality.

Claims 1, 20, 39, 58, 77 and 81-84 list computational steps in a program without tangible, useful, concrete result. The claims do not have any visible result or output. The steps of "determining" are missing real world result.

Claims 81-84 all state the intended use by use of word "useable". To overcome this type of rejection, claims could be amended to recite definite functionality (i.e. executed" or "processed")

Claims 82 and 84 are not limited to tangible embodiments. In view of Applicant's disclosure, specification page 34, line 23, the medium is not limited to tangible embodiments, in this case a carrier wave. As such, the claim is not limited to statutory subject matter and is therefore non-statutory.

To overcome this type of 101 rejection the claims need to be amended to include only the physical computer media and not a transmission media or other intangible or non-functional media. For this specification, carrier medium and transmission media would be not statutory but storage media would be statutory.

Claim Rejections - 35 USC § 102

12. The following is a quotation of the appropriate paragraphs of 35 U.S.C. 102 that form the basis for the rejections under this section made in this Office action:

> A person shall be entitled to a patent unless –
>
> (e) the invention was described in (1) an application for patent, published under section 122(b), by another filed in the United States before the invention by the applicant for patent or (2) a patent granted on an application for patent by another filed in the United States before the invention by the applicant for patent, except that an international application filed under the treaty defined in section 351(a) shall have the effects for purposes of this subsection of an application filed in the United States only if the international application designated the United States and was published under Article 21(2) of such treaty in the English language.

13. Claim 1-5, 9-10,14-24, 28-29, 33-43, 47-48, 52-62, 66-67, 71-76 and 81-88 are rejected under 35 U.S.C. 102(e) as being anticipated by Sundaresan et al. (U.S. Patent 6,606,620 B1).

As per claim 1 Sundaresan et al. is directed to a method of determining predictive models for a linked event detection system comprising the steps of:

determining source-identified training stories (column 3, lines 16-17, wherein "stories" means "documents");

determining inter-story similarity vectors for at least one story-pair (column 4, lines 23-24, wherein "similarity vectors" mean "similar terms");

determining link label information for the at least one story-pair (column 9, lines 8-9);

and determining at least one predictive model based on the inter-story similarity vector and the link label information (column 10, lines 5-13).

As per claim 2 Sundaresan et al. is directed to a step of determining inter-story similarity vectors comprises the steps of:

determining at least one inter-story similarity metric for the story-pairs (column 4, lines 9-25);

and determining at least one source-pair statistics for the at least one story-pair (column 10, lines 15-17).

As per claim 3 Sundaresan et al. is directed to a determining inter-story similarity vectors further comprise the step of normalizing the inter-story similarity metric based on the source-pair statistics (column 10, lines 17-22).

As per claim 4 Sundaresan et al. is directed to a determining inter-story similarity vectors further comprise the step of incrementally normalizing the inter-story similarity metric based on the source-pair statistics (column 10, lines 16-22).

As per claim 5 Sundaresan et al. is directed to the inter-story similarity metric is normalized based on at least one of subtraction and division (column 8, lines 22-27).

As per claim 9 Sundaresan et al. is directed to a comprising the step of transforming the source-identified training stories (column 1, line 63, wherein the "training stories" are in English).

As per claim 10 Sundaresan et al. is directed to transforming the source-identified training stories is at least one of translating, transcribing and linguistically transforming (column 1, line 63; column 2, line 43, wherein the HTML and XML are in English, therefore translation will not be necessary).

As per claim 14 Sundaresan et al. is directed to at least one inter-story similarity metric is normalized based on at least one of a source-pair identified similarity statistic (column 10, lines 15-17).

As per claim 15 Sundaresan et al. is directed to at least one predictive model is at least one of: a classifier, a support vector machine, a decision tree and a Naive-Bayes classifier (column 3, lines 13-14).

As per claim 16 Sundaresan et al. is directed to at least one of the source-pair similarity statistics are determined based on a source hierarchy (column 3, lines 50-51).

As per claim 17 Sundaresan et al. is directed to the source hierarchy is determined based on at least one source characteristic (column 3, lines 61-65, wherein "characteristic" means "leaf").

As per claim 18 Sundaresan et al. is directed to the source characteristic is at least one of a language characteristic, an input mode characteristic, a genre characteristic, a source name characteristic and a transformation characteristic (column 3, lines 54-60, wherein "language characteristic" means how the words in a document are related to each other).

As per claim 19 Sundaresan et al. is directed to the source-pair similarity statistic for a new source is determined based on at least one source characteristic of the new source (column 3, lines 50-53, wherein each new source has different hierarchy).

As per claim 20 Sundaresan et al. is directed to a linked event detection training system comprising:

an input/output circuit (column 7, lines 34-35, wherein it is inherent for computer to have input/output device circuit);

a memory (column 7, lines 34-35, wherein it is inherent for computer to have memory);

a processor that receives source-identified training stories and associated link label information for at least one story-pair via the input/output circuit (column 7, lines 34-35, wherein it is inherent for computer to have a processor);

an inter-story similarity vector determining circuit that determines an inter-story similarity vector for at least one story-pair (column 4, lines 23-24, wherein "similarity vectors" mean "similar terms");

and a predictive model determining circuit that determines at least one predictive model based on the inter-story similarity vector and the link label information (column 10, lines 5-13).

As per claim 21 Sundaresan et al. is directed to the inter-story similarity vector determining circuit is comprised of:

a similarity metric determining circuit that determines at least one inter-story similarity metric for the at least one story-pair (column 4, lines 9-25);

and a similarity statistics determining circuit that determines at least one source-pair statistic for the at least one story-pair (column 10, lines 15-17).

As per claim 22 Sundaresan et al. is directed to the inter-story similarity vector determining circuit normalizes the inter-story similarity metric based on the source-pair statistics (column 10, lines 17-22).

As per claim 23 Sundaresan et al. is directed to the inter-story similarity vector determining circuit incrementally normalizes the inter-story similarity metric based on the source-pair statistics (column 10, lines 16-22).

As per claim 24 Sundaresan et al. is directed to at least one of the inter-story similarity metrics is normalized based on at least one of a subtraction and a division operation (column 8, lines 22-27).

As per claim 28 Sundaresan et al. is directed to a comprising the step of transforming the source-identified training stories (column 1, line 63, wherein the "training stories" are in English).

As per claim 29 Sundaresan et al. is directed to transforming the source-identified training stories is at least one of translating, transcribing and linguistically transforming (column 1, line 63; column 2, line 43, wherein the HTML and XML are in English therefore translation will not be necessary).

As per claim 33 Sundaresan et al. is directed to the at least one inter-story similarity metric is normalized based on at least one of a source-pair identified similarity statistic (column 10, lines 15-17).

As per claim 34 Sundaresan et al. is directed to the at least one predictive model is at least one of: a classifier, a support vector machine, a decision tree and a Naive-Bayes classifier (column 3, lines 13-14).

As per claim 35 Sundaresan et al. is directed to the source-pair identified similarity statistic is determined based on a source hierarchy (column 3, lines 50-51).

As per claim 36 Sundaresan et al. is directed to the source hierarchy is determined based on at least one of a source characteristic (column 3, lines 61-65, wherein "characteristic" means "leaf").

As per claim 37 Sundaresan et al. is directed to the source characteristic is at least one of a language characteristic, an input mode characteristic, a genre

characteristic, a source name characteristic and a transformation characteristic (column 3, lines 54-60, wherein "language characteristic" means how the words in a document are related to each other).

As per claim 38 Sundaresan et al. is directed to the source-pair similarity statistic for a new source is determined based on at least one source characteristics of the new source (column 3, lines 50-53, wherein each new source has different hierarchy).

As per claim 39 Sundaresan et al. is directed to a method of linked event detection comprising the steps of:

determining source-identified training stories (column 3, lines 16-17, wherein "stories" means "documents");

determining inter-story similarity vectors for the story-pairs (column 4, lines 23-24, wherein "similarity vectors" mean "similar terms");

determining at least one predictive model for link detection (column 10, lines 5-13);

and determining a link between the story-pairs based on the predictive model and the inter-story similarity vector (column 10, lines 5-13, wherein sorting determines the link).

As per claim 40 Sundaresan et al. is directed to a step of determining inter-story similarity vectors comprises the steps of:

determining at least one inter-story similarity metric for each story-pair (column 4, lines 9-25);

and determining source-pair statistics for the story-pairs (column 10, lines 15-17).

As per claim 41 Sundaresan et al. is directed to a determining inter-story similarity vectors further comprise the step of normalizing the inter-story similarity metric based on the source-pair statistics (column 10, lines 17-22).

As per claim 42 Sundaresan et al. is directed to a determining inter-story similarity vectors further comprise the step of incrementally normalizing the inter-story similarity metric based on the source-pair statistics (column 10, lines 16-22).

As per claim 43 Sundaresan et al. is directed to the inter-story similarity metric is normalized based on at least one of subtraction and division (column 8, lines 22-27).

As per claim 47 Sundaresan et al. is directed to a comprising the step of transforming the source-identified training stories (column 1, line 63, wherein the "training stories" are in English).

As per claim 48 Sundaresan et al. is directed to transforming the source-identified training stories is at least one of translating, transcribing and linguistically

transforming (column 1, line 63; column 2, line 43, wherein the HTML and XML are in English therefore translation will not be necessary).

As per claim 52 Sundaresan et al. is directed to the at least one inter-story similarity metric is normalized based on at least one of a source-pair identified similarity statistic. (column 10, lines 15-17).

As per claim 53 Sundaresan et al. is directed to the at least one predictive model is at least one of: a classifier, a support vector machine and a decision tree, a Naive-Bayes-classifier (column 8, lines 22-27).

As per claim 54 Sundaresan et al. is directed to the source-pair similarity statistic for a new source is determined based on at least one source characteristics of the new source (column 3, lines 50-51).

As per claim 55 Sundaresan et al. is directed to the source hierarchy is determined based on at least one of a source characteristic (column 3, lines 61-65, wherein "characteristic" means "leaf").

As per claim 56 Sundaresan et al. is directed to the source characteristic is at least one of a language characteristic, an input mode characteristic, a genre characteristic, a source name characteristic and a transformation characteristic (column

3, lines 54-60, wherein "language characteristic" means how the words in a document are related to each other).

As per claim 57 Sundaresan et al. is directed to the source-pair similarity statistic for a new source is determined based on at least one source characteristics of the new source (column 3, lines 50-53, wherein each new source has different hierarchy).

As per claim 58 Sundaresan et al. is directed to linked event detection system comprising:

an input/output circuit (column 7, lines 34-35, wherein it is inherent for computer to have input/output device circuit);

a memory (column 7, lines 34-35, wherein it is inherent for computer to have memory);

a processor that receives source-identified training stories via the input/output circuit (column 7, lines 34-35, wherein it is inherent for computer to have processor);

an inter-story similarity vector determining circuit that determines inter-story similarity vectors for the story-pairs (column 4, lines 23-24, wherein "similarity vectors" mean "similar terms");

and a link determining circuit that determines links between story-pairs based on a predictive model and the inter-story similarity vectors (column 10, lines 5-13, wherein sorting determines the link).

As per claim 59 Sundaresan et al. is directed to the inter-story similarity vector determining circuit is comprised of:

a similarity metric determining circuit that determines at least one inter-story similarity metric for the story-pairs (column 4, lines 9-25);

and a similarity statistics determining circuit that determines source-pair statistics for the story-pairs (column 10, lines 15-17).

As per claim 60 Sundaresan et al. is directed to the inter-story similarity vector determining circuit normalizes the inter-story similarity metric based on the source-pair statistics (column 10, lines 17-22).

As per claim 61 Sundaresan et al. is directed to the inter-story similarity vector determining circuit incrementally normalizes the inter-story similarity metric based on the source-pair statistics (column 10, lines 16-22).

As per claim 62 Sundaresan et al. is directed to at least one of the inter-story similarity metrics is normalized based on at least one of a subtraction and a division operation (column 8, lines 22-27).

As per claim 66 Sundaresan et al. is directed to a comprising the step of transforming the source-identified training stories (column 1, line 63, wherein the "training stories" are in English).

As per claim 67 Sundaresan et al. is directed to transforming the source-identified training stories is at least one of translating, transcribing and linguistically transforming (column 1, line 63; column 2, line 43, wherein the HTML and XML are in English therefore translation will not be necessary).

As per claim 71 Sundaresan et al. is directed to the at least one inter-story similarity metric is normalized based on at least one of a source-pair identified similarity statistic (column 10, lines 15-17).

As per claim 72 Sundaresan et al. is directed to the predictive model is at least one of: a classifier, a support vector machine and a decision tree, a Naive-Bayes classifier (column 8, lines 22-27).

As per claim 73 Sundaresan et al. is directed to the source-pair identified similarity statistic is determined based on a source hierarchy (column 3, lines 50-51).

As per claim 74 Sundaresan et al. is directed to the source hierarchy is determined based on at least one of a source characteristic (column 3, lines 61-65, wherein "characteristic" means "leaf").

As per claim 75 Sundaresan et al. is directed to the source characteristic is at least one of a language characteristic, an input mode characteristic, a genre characteristic, a source name characteristic and a transformation characteristic (column 3, lines 54-60, wherein "language characteristic" means how the words in a document are related to each other).

As per claim 76 Sundaresan et al. is directed to the source-pair similarity statistic for a new source is determined based on at least one source characteristics of the new source (column 3, lines 50-53, wherein each new source has different hierarchy).

As per claim 81 Sundaresan et al. is directed to computer readable storage medium comprising: computer readable program code embodied on the computer readable storage medium, the computer readable program code usable to program a computer to determine at least one predictive model for a linked event detection system comprising the steps of:

determining source-identified training stories (column 3, lines 16-17, wherein "stories" means "documents");

determining inter-story similarity vectors for at least one story-pair (column 4, lines 23-24, wherein "similarity vectors" mean "similar terms");

determining link label information for the at least one story-pair (column 9, lines 8-9);

and determining at least one predictive model based on the inter-story similarity vector and the link label information (column 10, lines 5-13).

As per claim 82 Sundaresan et al. is directed to a carrier wave encoded to transmit a control program, useable to program a computer to determine a predictive model for a linked event detection system, to a device for executing the program, the control program comprising:

instructions for determining source-identified training stories (column 3, lines 16-17, wherein "stories" means "documents");

instructions for determining inter-story similarity vectors for at least one story-pair (column 4, lines 23-24, wherein "similarity vectors" mean "similar terms");

instructions for determining link label information for the at least one story-pair (column 9, lines 8-9);

and instructions for determining at least one predictive model based on the inter-story similarity vector and the link label information (column 10, lines 5-13).

As per claim 83 Sundaresan et al. is directed to computer readable storage medium comprising: computer readable program code embodied on the computer readable storage medium, the computer readable program code usable to program a computer to detect linked events comprising the steps of:

determining source-identified training stories (column 3, lines 16-17, wherein "stories" means "documents");

determining inter-story similarity vectors for the the at least one story-pair (column 4, lines 23-24, wherein "similarity vectors" mean "similar terms");

determining at least one predictive model for link detection (column 9, lines 8-9);

determining a link between story-pairs based on the at least one predictive model and the inter-story similarity vector (column 10, lines 5-13).

As per claim 84 Sundaresan et al. is directed to a carrier wave encoded to transmit a control program, useable to program a computer to detect linked events, to a device for executing the program, the control program comprising:

instructions for determining source-identified training stories (column 3, lines 16-17, wherein "stories" means "documents");

instructions for determining inter-story similarity vectors for the at least one story-pair (column 4, lines 23-24, wherein "similarity vectors" mean "similar terms");

instructions for determining at least one predictive model for link detection (column 9, lines 8-9);

instructions for determining a link between story-pairs based on the predictive model and the inter-story similarity vector (column 10, lines 5-13).

As per claims 85 and 86 Sundaresan et al. is directed to determining at least one source-pair statistic for the at least one story-pair is based on at least one of a similarity metric and a statistic associated with the metric (column 3, lines 25-29, wherein the statistical algorithm uses metric for the computations).

As per claims 87 and 88 Sundaresan et al. is directed to at least one of the predictive models is a trained predictive model (column 10, lines 29-33, wherein the "trained predictive model" is determined by use of statistical model).

14. The following is a quotation of the appropriate paragraphs of 35 U.S.C. 102 that form the basis for the rejections under this section made in this Office action:

> A person shall be entitled to a patent unless –
>
> (a) the invention was known or used by others in this country, or patented or described in a printed publication in this or a foreign country, before the invention thereof by the applicant for a patent.

15. Claim 77-80 are rejected under 35 U.S.C. 102(a) as being anticipated by Brown, Ralf D. "Dynamic Stopwording for Story Link Detection", (hereafter referred as Brown).

As per claim 77 Brown is directed to a method of determining a stopword list comprising the steps of:

determining a source-identified training corpus of text information (page 1, column 2, lines 26-29);

determining a verified first transformation of the source-identified training corpus text from a first source mode to a second source mode (page 1, column 2, lines 26-29; page 1 column 2, lines 33-40, wherein the "transformation" would be the "single-pass incremental clustering method");

determining an un-verified second transformation of the source-identified training corpus text from a first source mode to a second source mode (page 1, column 2, lines 17-18, wherein "un-verified" means any "story from a newswire");

determining at least one transformation errors associated with distribution differences between the first and second transformations and identified sources (page 2, column 2, lines 4-6);

determining at least one source-specific transformation actions for the determined transformation errors (page 2, column 1, lines 2-6).

As per claim 78 Brown is directed to the first source mode is at least one of a text source, an optical character recognition source and an automatic speech recognition source (page 1, column 2, lines 22-24).

As per claim 79 Brown is directed to the second source mode is at least one of a text source, an optical character recognition source and an automatic speech recognition source (page 1, column 2, lines 22-24; page 2, column 1, lines 6-8).

As per claim 80 Brown is directed to wherein the source-specific transformation is at least one of a removal, a repair and a normalization transformation (page 2, column 1, lines 4-6).

Claim Rejections - 35 USC § 103

16. The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

> (a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negatived by the manner in which the invention was made.

17. Claims 6-8, 25-27, 44-46, and 63-65 are rejected under 35 U.S.C. 103(a) as being unpatentable over Sundaresan et al. (US Patent 6,606,620 B1) in view of Gange et al. (US 2004/006559 A1).

As per claims 6, 25, 44 and 63 Sundaresan et al. fails to teach the use of probability based metric and a Euclidean based similarity metric.

Gange et al. teaches the use of Euclidean distance (page 3, paragraph 0045).

It would have been obvious to one in of ordinary skill in the art at the time the invention was made to modify Sundaresan et al. by teachings of Gange et al. to include the use of Euclidean distance as it is metrics often used in the database field to compute distances between similar terms.

As per claims 7, 26, 45 and 64 Sundaresan et al. as modified fails to teach the use of similarity metric is at least one of a Hellinger, a Tanimoto and a clarity distance based metric.

Gange et al. teaches the use of Tanimoto coefficient (page 3, paragraph 0045).

It would have been obvious to one in of ordinary skill in the art at the time the invention was made to modify Sundaresan et al. by teachings of Gange et al. to include the use of Tanimoto coefficient as it is metrics often used in the database field to compute distances between similar terms.

A per claims 8, 27, 46 and 65 Sundaresan et al. as modified fails to teach the use of inter-story similarity metric is a cosine-distance based metric.

Gange et al. teaches the use of Cosine coefficient (page 3, paragraph 0045).

It would have been obvious to one in of ordinary skill in the art at the time the invention was made to modify Sundaresan et al. by teachings of Gange et al. to include the use of Cosine coefficient as it is metrics often used in the database field to compute distances between similar terms

18. Claims 11-13, 30-32, 49-51 and 68-70 are rejected under 35 U.S.C. 103(a) as being unpatentable over Sundaresan et al. (US Patent 6,606,620 B1) in view of Zhou (US 2004/0002849 A1).

As per claims 11, 30, 49 and 68 Sundaresan et al. fails to teach the inter-story similarity metrics are based on terms in at least one source-identified term frequency-inverse story frequency models.

Zhou teaches the use of frequency-inverse (page 3, column 2, paragraph 0030, lines 9-11).

It would have been obvious to one in of ordinary skill in the art at the time the invention was made to modify Sundaresan et al. by teachings of Zhou to include the use of frequency-inverse because it predicts effective example of sentence retrieval as stated on page 1, column 1, paragraph 0005 of Zhou.

As per claims 12, 37, 50 and 69 Sundaresan et al. as modified fails to teach the terms in source-identified term frequency-inverse story frequency models are based on language.

Zhou teaches that the retrieved samples are to aid in writing or translation (page 3, paragraph 0030, lines 2-4, wherein writing or translating has basis in language).

It would have been obvious to one in of ordinary skill in the art at the time the invention was made to modify Sundaresan et al. by teachings of Zhou to include the inverse-frequency based on language because term comparison includes terms of a language.

As per claims 13, 32, 51 and 70 Sundaresan et al. as modified fails to teach determining terms comprises the steps: determining a reference language; and determining reference language and non-reference language terms.

Zhou teaches the changing of sample terms from one mode to another (page 3, paragraph 0032).

It would have been obvious to on of ordinary skill in the art at the time the invention was made to modify Sundaresan et al. by teachings of Zhou to include the

determination of reference language since the correct translation requires the correct reference language.

Conclusion

19. Any inquiry concerning this communication or earlier communications from the examiner should be directed to Tomasz Ponikiewski whose telephone number is (571)272-1721. The examiner can normally be reached on 8:00-4:30.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Jeffrey A. Gaffin can be reached on (571)272-4146. The fax phone number for the organization where this application or proceeding is assigned is 571-273-8300.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see http://pair-direct.uspto.gov. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free).

Tomasz Ponikiewski
February 24, 2006

JEFFREY GAFFIN
SUPERVISORY PATENT EXAMINER
TECHNOLOGY CENTER 2100